



Mini Projeto – Walmart

1. Objetivo do Negócio

Nossa equipe, contratada pela empresa Walmart, realizará uma consultoria estratégica que, por meio de uma análise crítica e aprofundada no setor comercial das diversas lojas da empresa, recomendará a melhor loja para expandir seu tamanho.

Analisaremos os dados disponibilizados, colhendo as informações relevantes e, ao final, auferindo os insights, recomendaremos uma loja que apresentará o melhor potencial financeiro para sua expansão.

Para o melhor entendimento e esclarecimento das indicações serão expostos com precisão as métricas utilizadas para o desenvolvimento dos modelos de análise e da referida recomendação, ressaltando as passagens mais complexas e insights obtidos durante a resolução do projeto.

2. Business Understanding

2.1. Apresentação

A Walmart Inc. é uma multinacional estadunidense de lojas de departamento. A companhia foi fundada por Sam Walton em 1962, incorporada em 31 de outubro de 1969 e abriu seu capital na New York Stock Exchange (NYSE) em 1972.

Para se ter ideia, no ano de 2021, a companhia obteve um lucro de \$13.51 Bilhões de dólares, sendo uma das principais lojas de varejo do mundo.

A Walmart realiza vários eventos promocionais de descontos ao longo do ano e essas remarcações precedem feriados importantes. Os quatro maiores de todos incluem o Super Bowl, o Dia do Trabalho, o Dia de Ação de Graças e o Natal, sendo que, as semanas referentes a esses feriados têm um peso maior.

Os dados disponibilizados para a análise contemplam as vendas semanais de 45 lojas espalhadas nos Estados Unidos.

2.2. Dados

Os dados utilizados foram extraídos da seguinte fonte:

<https://www.kaggle.com/datasets/yasserh/walmart-dataset>

Numa primeira análise é possível verificar que os dados apresentam as vendas semanais das 45 lojas da rede Walmart, além de outras métricas econômicas e meteorológicas no período entre 05/02/2010 e 01/11/2012.

2.3. Dicionários dos dados

As features disponíveis no dataset apresentam um total de 8 atributos e 6435 registros.:

- **Store** – Número/Identificador único da loja
- **Date** – Semana de venda
- **Weekly_Sales** – Vendas semanais referentes a loja específica
- **Holiday_Flag** – 1 Quando é uma semana com feriado
- **Holiday_Flag** – 0 Quando não é uma semana com feriado
- **Temperature** – Temperatura no dia da venda em Fahrenheit [°F] (Posteriormente transformada para Graus Celsius [°C])
- **Fuel_Price** – Preço do combustível na região em dólar [\$]

- **CPI – (Costumer Price Index) Índice de preço ao consumidor**
- **Unemployment – Taxa de Desemprego**
- **Holiday Events(Eventos de Feriados):**
Super Bowl: 12/02/2010, 11/02/2011, 10/02/2012, 08/02/2012
Labour Day: 10/09/2010, 09/09/2011, 07/09/2012, 06/09/2013
Thanksgiving: 26/11/2010, 25/11/2011, 23/11/2012, 29/11/2013
Christmas: 31/12/2010, 30/12/2011, 28/12/2012, 27/12/2013

Além dos registros listados acima, foram criados os demais atributos para análise:

- **Month – Mês da venda**
- **Year – Ano da venda**
- **unemployment_media – Média da taxa de desemprego por loja**
- **Fuel_Price_media – Média do valor do combustível**
- **weekly_sales_media – Média do valor vendido por semana por loja**

3. Coleta de Dados

```
[5] # Instalando as bibliotecas
```

```
!pip install pandas
!pip install numpy
!pip install matplotlib
!pip install seaborn
```

```
[6] # Importando as bibliotecas
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[7] # Baixando e carregando o Dataset
```

```
url = "https://raw.githubusercontent.com/Caricchio-Data/DataExpertClassDNC/main/Walmart.csv"
table = pd.read_csv(url)
df = pd.DataFrame(table)
```

```
[8] # Verificando as informações do dataset
df.info()
```

```
[9] # Verificando as dimensões do dataset
df.shape
```

4. Limpeza dos Dados

```
[26] # Verificando dados faltantes

df.isnull().sum()
```

```
[27] # Verificando valores duplicados

df.duplicated().sum()
```

```
[12] # Alterando a coluna Temperature para °C (Celsius)

df['Temperature'] = df.Temperature.apply(lambda x : (x - 32) * (5 / 9))
df
```

```
[13] # Ajustando o formato de data

df['Date'] = pd.to_datetime(df['Date'], dayfirst = True)
df['Month'] = df['Date'].dt.month
df['Year'] = df['Date'].dt.year
df.head(10)
```

```
[14] # Validação de valores de elementos em tabela
df.nunique().sort_values()
```

```
[15] # Estatística descritiva das variáveis numéricas
pd.set_option('display.float_format', lambda x: '%.2f' % x)

# Ajustando a exibição das colunas para análise da Estatística descritiva
colunas_numericas = ['Weekly_Sales', 'Temperature', 'Fuel_Price', 'CPI', 'Unemployment']
df[colunas_numericas].describe().T
```

5. Exploração dos Dados

- Quais serão as métricas essenciais utilizadas?
- Com base nos dados, qual loja deverá ser expandida pela Walmart?

Dentre os atributos simples, compostos, derivados e chave, foram analisados o faturamento médio geral, média de vendas por período, número da loja, temperatura, índice de preços ao consumidos, taxa de desemprego e *Holiday flags*.

As métricas iniciais definidas visam verificar a lucratividade, análise de crescimento e correlação entre os demais atributos com o fim de exibir negócios com maior probabilidade de retorno financeiro após a expansão.

Para o aprofundamento da análise, os demais fatores foram analisados como forma de assegurar sua irrelevância no exame dos insights.

Inicialmente, foram verificadas as lojas com maiores médias de vendas no período analisado.

```
[16] # Lojas com as maiores média de venda no período analisado

stores_sales_mean = ((df.groupby(['Store']).mean()).sort_values(by=['Weekly_Sales'], ascending=False)['Weekly_Sales'])
stores_sales_mean

# Resultado da análise: As lojas 20 , 4, 14, 13, 2 apresentaram o maior faturamento médio no período
```

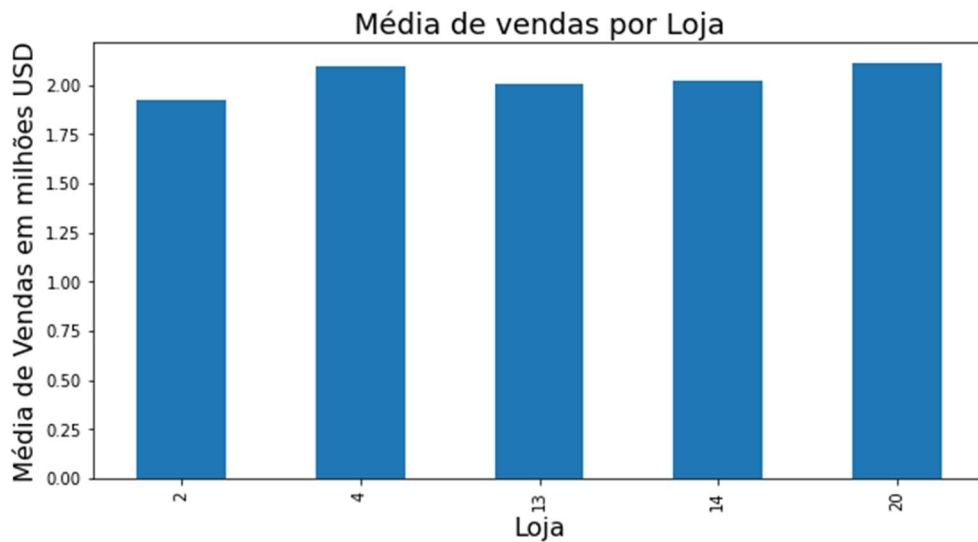
Em seguida, com base no resultado das maiores médias de vendas, para uma melhor visualização, foram expostos os dados das cinco lojas com maior faturamento médio e o respectivo gráfico da referida média de vendas em milhões.

```
[17] # Dados das lojas com o maior faturamento médio

sales_bigger = [ 20 , 4, 14, 13, 2]
stores_bigger_sales = df.query('Store in @sales_bigger')
stores_bigger_sales
```

```
[18] # Gráfico da média de venda por loja em milhões (5 Lojas com maior faturamento médio)

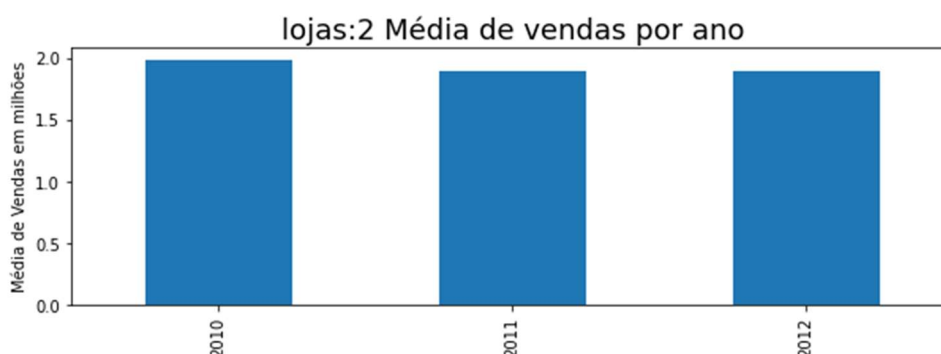
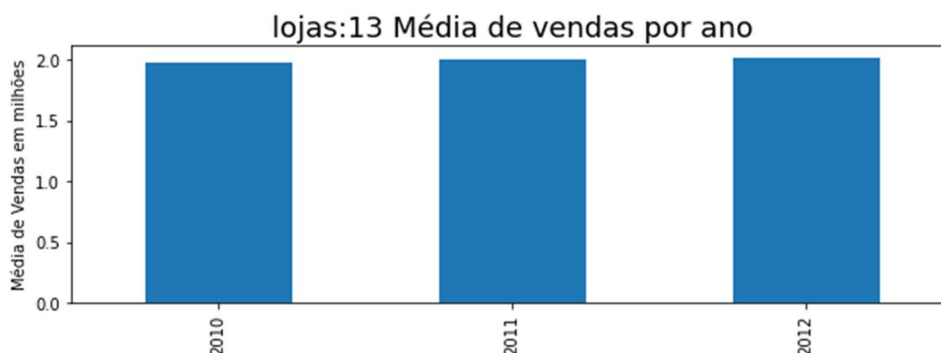
weekly_sales = stores_bigger_sales['Weekly_Sales'].groupby(df['Store']).mean() / 1000000
plt.figure(figsize=(10,5))
weekly_sales.plot.bar();
plt.title('Média de vendas por Loja', fontsize=18);
plt.ylabel('Média de Vendas em milhões USD', fontsize=16);
plt.xlabel('Loja', fontsize=16);
```

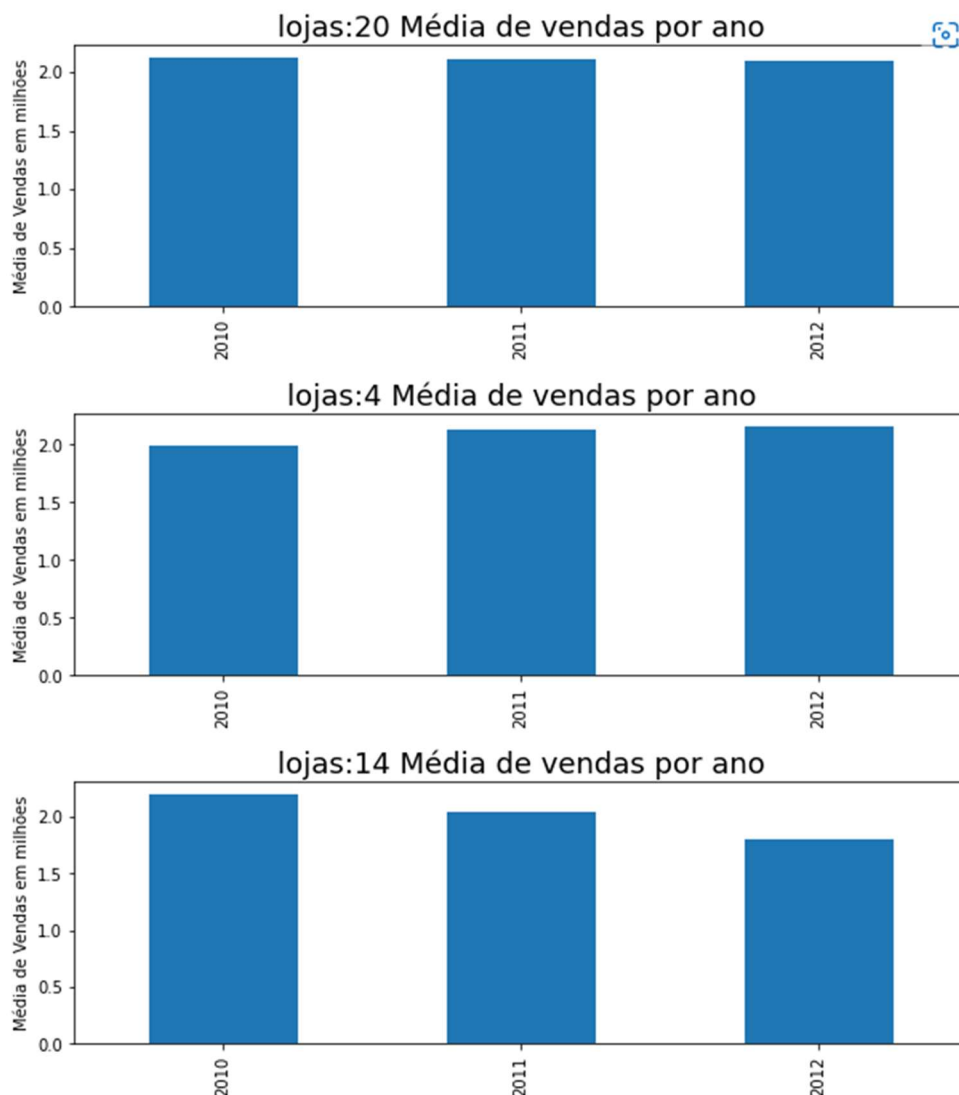


A partir das análises supra, foi possível indicar as lojas 4 e 20 se destacam dentre as 5 com maior faturamento médio.

Para examinar o crescimento ou diminuição no faturamento das 05(cinco) lojas com maior faturamento médio, foi realizado um gráfico, demonstrando o faturamento por ano de cada loja individualmente.

```
[19] # Gráfico da média de venda/mês de todos os anos (5 Lojas com maior faturamento médio)
for x in sales_bigger:
    stores_bigger_sales = df.query('Store == @x')
    weekly_sales = stores_bigger_sales['Weekly_Sales'].groupby(df['Year']).mean() / 1000000
    plt.figure(figsize=(10,3))
    weekly_sales.plot.bar();
    plt.title('lojas:%i Média de vendas por ano'%x, fontsize=18);
    plt.ylabel('Média de Vendas em milhões', fontsize=10);
    plt.xlabel('')
```





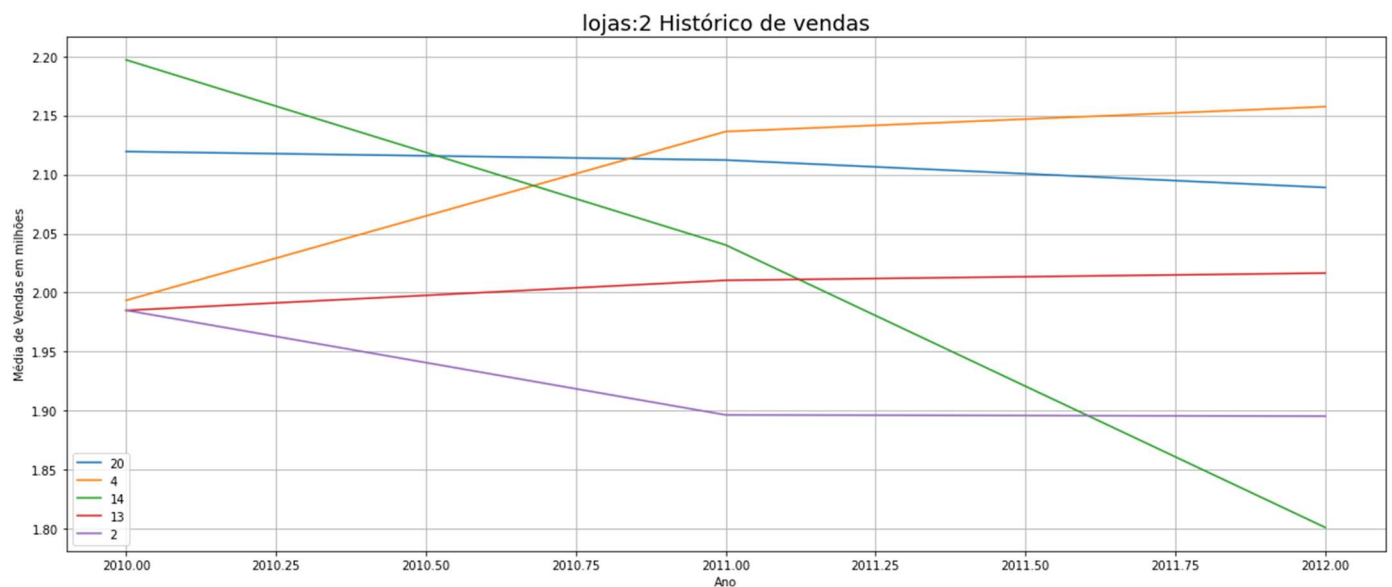
Numa breve análise dos dados expostos, verificando o crescimento ou diminuição anual no faturamento dentre as 05(cinco) lojas com maior faturamento, é possível identificar um importante fator de destaque para a escolha da loja objeto de investimento e expansão. O indicador de crescimento ou diminuição dentre lojas com alto faturamento se apresenta como um dos principais parâmetros iniciais para escolha do melhor negócio com receptividade para uma possível expansão.

Como confirmação e comparação clara dos dados obtidos, foi feito o gráfico comparativo de vendas por loja, entre as 5 lojas de maior faturamento, durante todo o período analisado.

```
[21] # Comparativo de vendas x anos entre as lojas (Lojas 20 , 4, 14, 13, 2 )
      for x in sales_bigger:

          stores_bigger_sales = df.query('Store == @x')
          weekly_sales = stores_bigger_sales['Weekly_Sales'].groupby(df['Year']).mean() / 1000000
          weekly_sales.plot(figsize=(20,8));
          plt.title('lojas:%i Histórico de vendas'%x, fontsize=18);
          plt.ylabel('Média de Vendas em milhões', fontsize=10);
          plt.xlabel('Ano')
          plt.grid(True)
          plt.legend(sales_bigger)

      plt.show()
```



A partir do gráfico supra é possível perceber que as Lojas 4 e 13 apresentaram crescimento, com destaque para a Loja 4 que demonstrou considerável e constante aumento nas vendas.

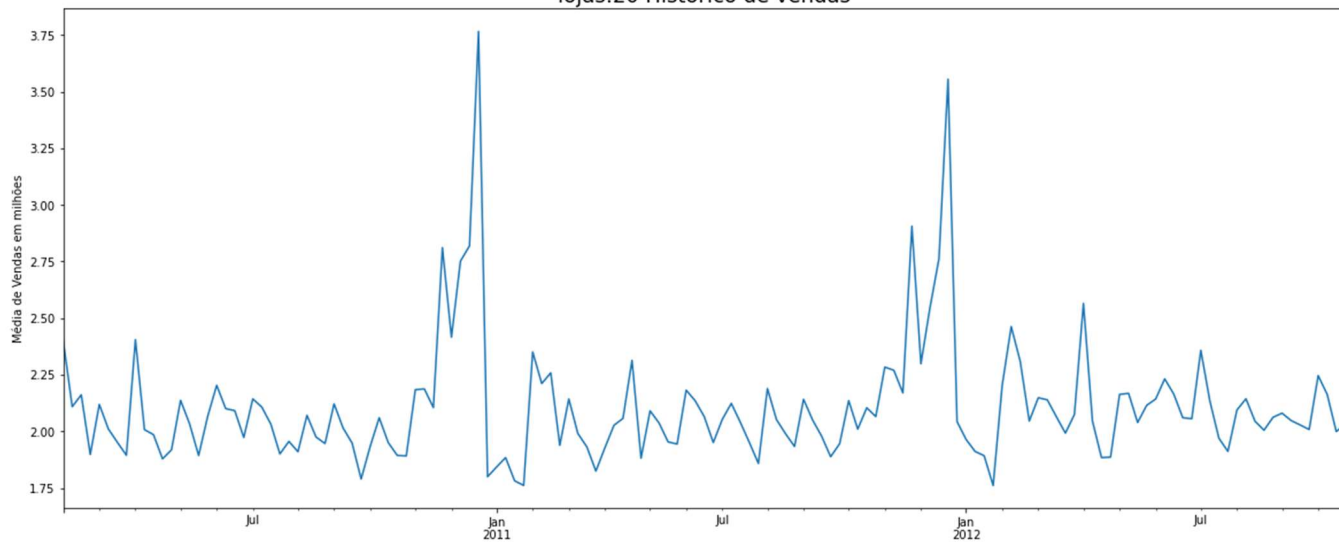
A loja 20 exibe leve queda ao longo dos anos a partir da metade do ano de 2010, demonstrando, entretanto, tendencia de permanecer em declínio.

As lojas 2 e 14 se encontram em constante involução, com destaque negativo para a Loja 14 que apresenta intensa desvalorização em todo o período analisado.

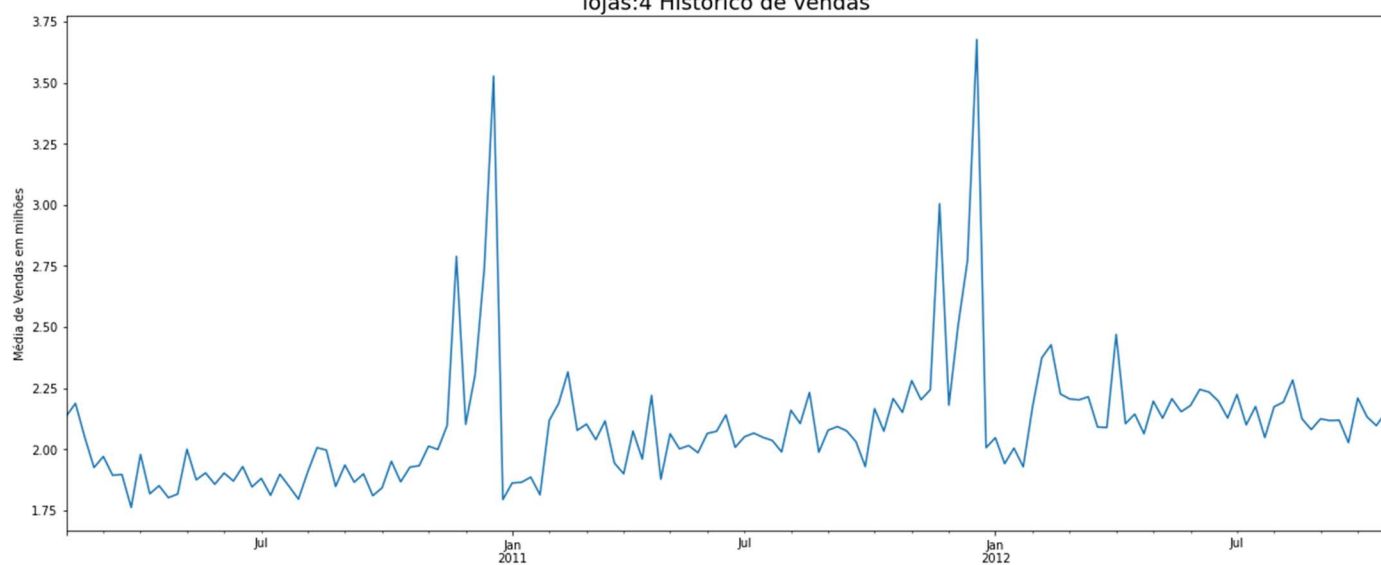
Em seguida, para identificar os efeitos dos principais feriados nas vendas, além de uma análise mais profunda do volume de vendas ao longo do ano, foram executados gráficos das 5(cinco) lojas com o histórico de vendas ao longo de todo o período analisado.

```
[20] # Histórico de vendas de todo o período
for x in sales_bigger:
    stores_bigger_sales = df.query('Store == @x')
    weekly_sales = stores_bigger_sales['Weekly_Sales'].groupby(df['Date']).mean() / 1000000
    plt.figure(figsize=(20,8))
    weekly_sales.plot();
    plt.title('lojas:%i Histórico de vendas'%x, fontsize=18);
    plt.ylabel('Média de Vendas em milhões', fontsize=10);
    plt.xlabel('')
```

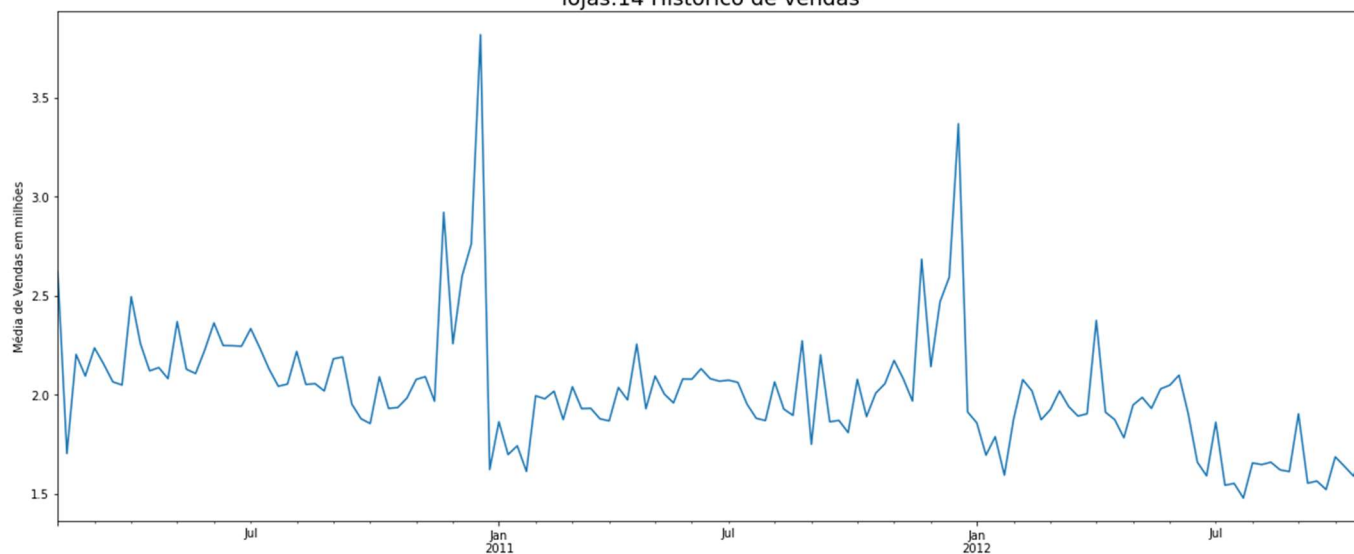

lojas:20 Histórico de vendas

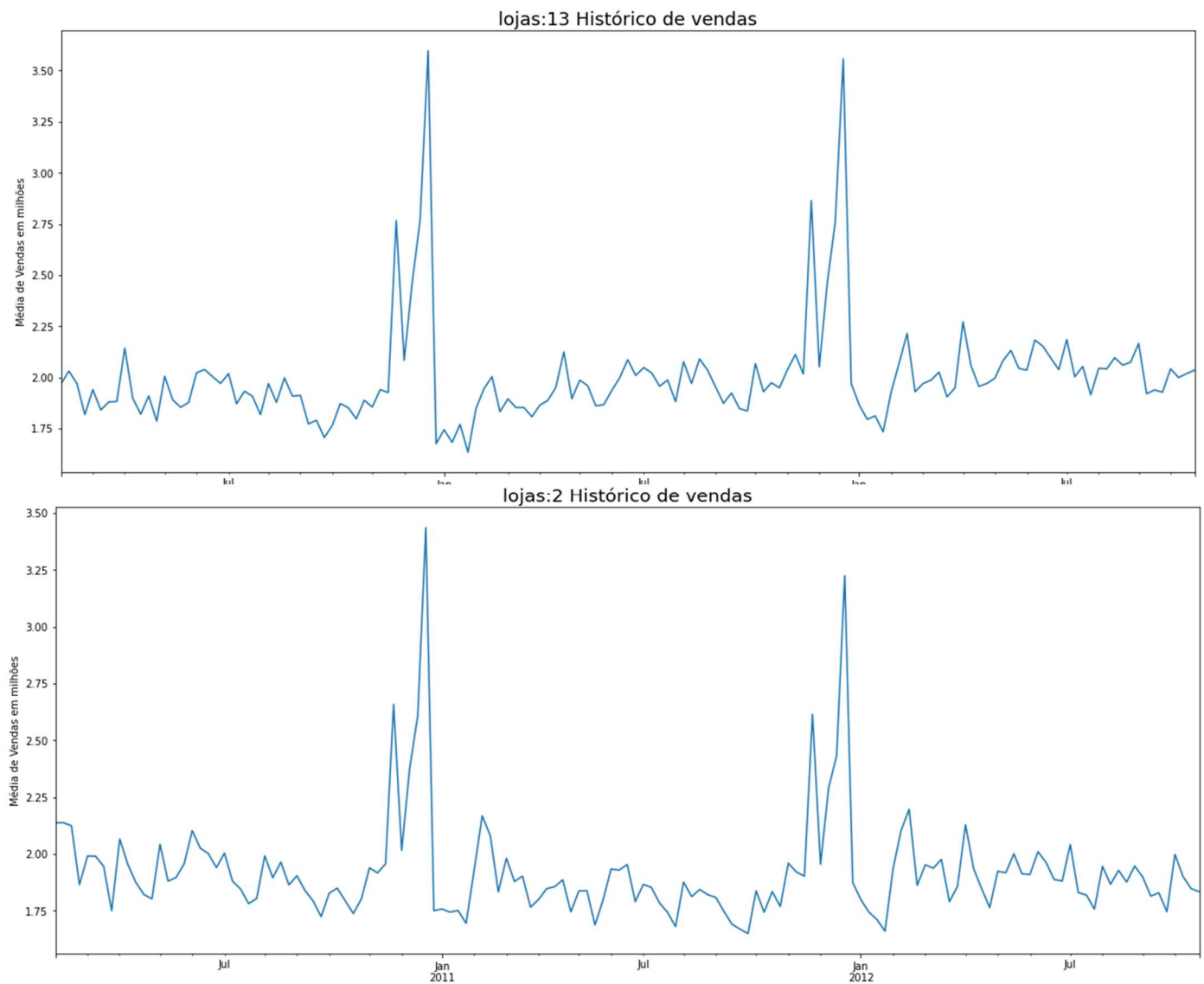


lojas:4 Histórico de vendas



lojas:14 Histórico de vendas





Observando os gráficos com os históricos de vendas ao longo do ano, percebe-se uma constante e significativa alta no período entre novembro e dezembro, época reconhecida pelos feriados de *Thanksgiving* e *Christmas*, que se repete em todos os anos analisados.

Ainda, no intuito de verificar os reflexos dos demais atributos no faturamento e crescimento de cada uma das 5(cinco) lojas com maior faturamento médio, foram realizados gráficos com os históricos de temperaturas, preços do combustível, índice de preços ao consumidor e taxa de desemprego.

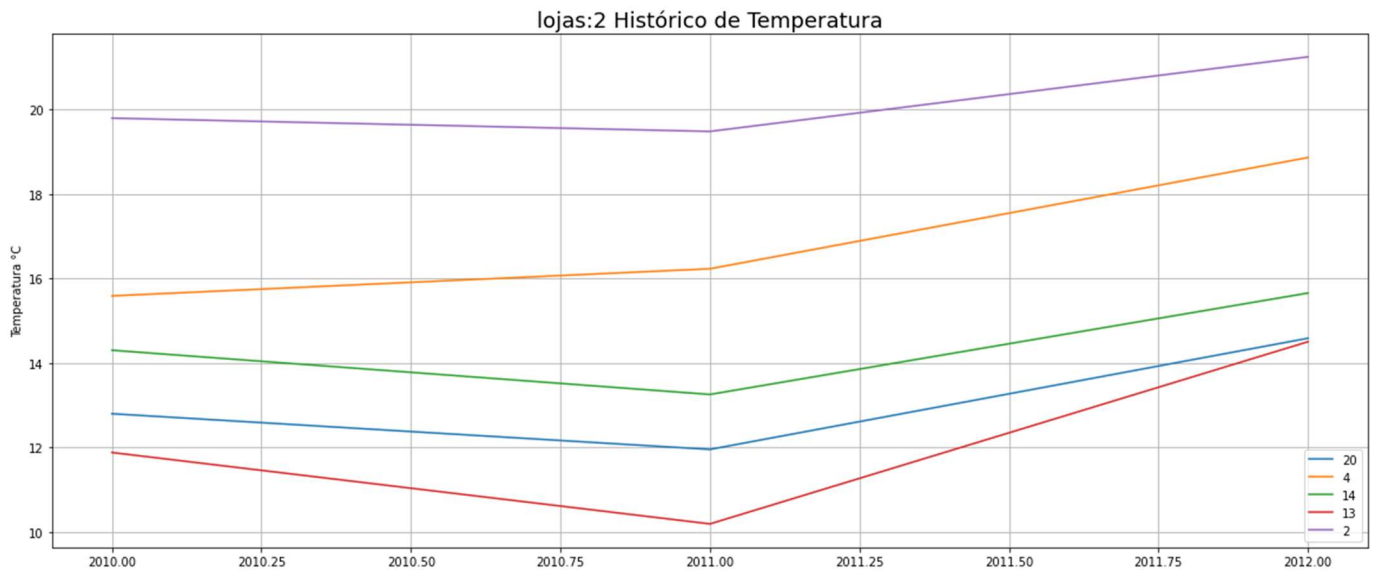
Quanto a análise histórica do clima pela temperatura:

```
[22] # Análise do histórico de temperaturas (Lojas 20 , 4, 14, 13, 2 )
```

```
for x in sales_bigger:

    stores_bigger_sales = df.query('Store == @x')
    weekly_sales = stores_bigger_sales['Temperature'].groupby(df['Year']).mean()
    weekly_sales.plot(figsize=(20,8));
    plt.title('lojas:%i Histórico de Temperatura'%x, fontsize=18);
    plt.ylabel('Temperatura °C', fontsize=10);
    plt.xlabel('')
    plt.grid(True)
    plt.legend(sales_bigger)

plt.show()
```



Não é possível identificar nenhum fator de destaque quanto as temperaturas das lojas analisadas que tenha a consistência necessária para alterar o potencial financeiro destas. Em nenhum dos casos houve a constatação de calor ou frio extremos.

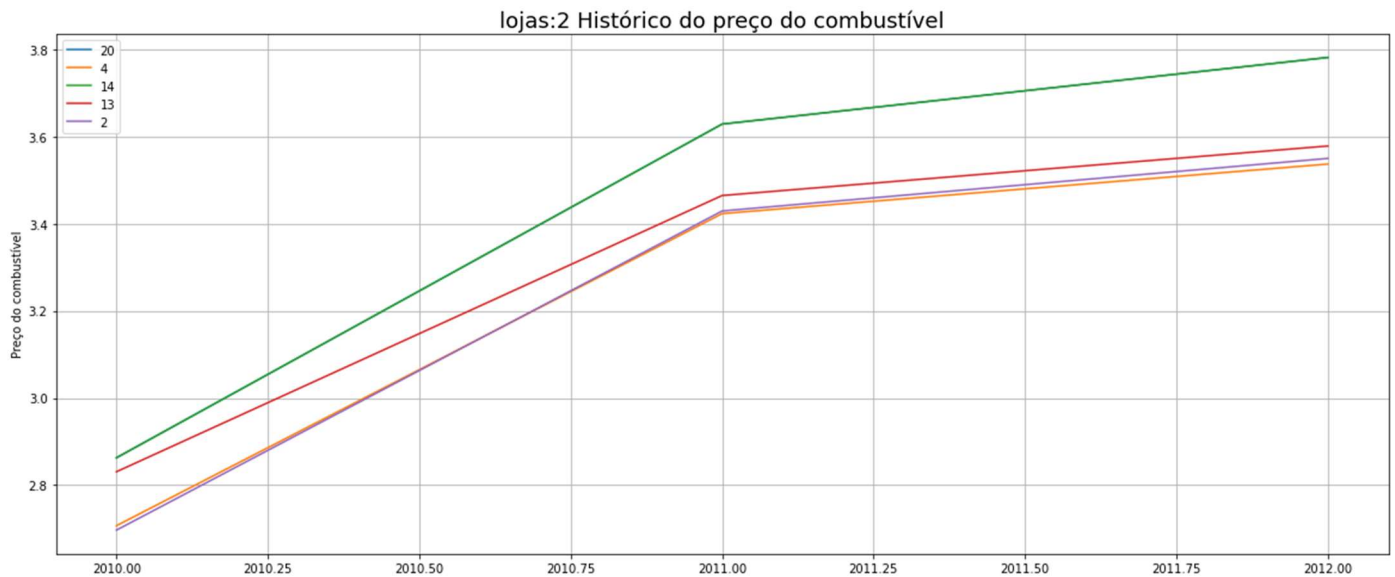
Quanto a variação histórica do preço do combustível:

```
[23] # Análise do histórico do preço do combustível (Lojas 20 , 4, 14, 13, 2 )
```

```
for x in sales_bigger:

    stores_bigger_sales = df.query('Store == @x')
    weekly_sales = stores_bigger_sales['Fuel_Price'].groupby(df['Year']).mean()
    weekly_sales.plot(figsize=(20,8));
    plt.title('lojas:%i Histórico do preço do combustível'%x, fontsize=18);
    plt.ylabel('Preço do combustível', fontsize=10);
    plt.xlabel('')
    plt.grid(True)
    plt.legend(sales_bigger)

plt.show()
```



Também não foram identificados fatores relevantes quanto a variação do preço do combustível entre as lojas analisadas que reflitam diretamente na lucratividade e crescimento das lojas.

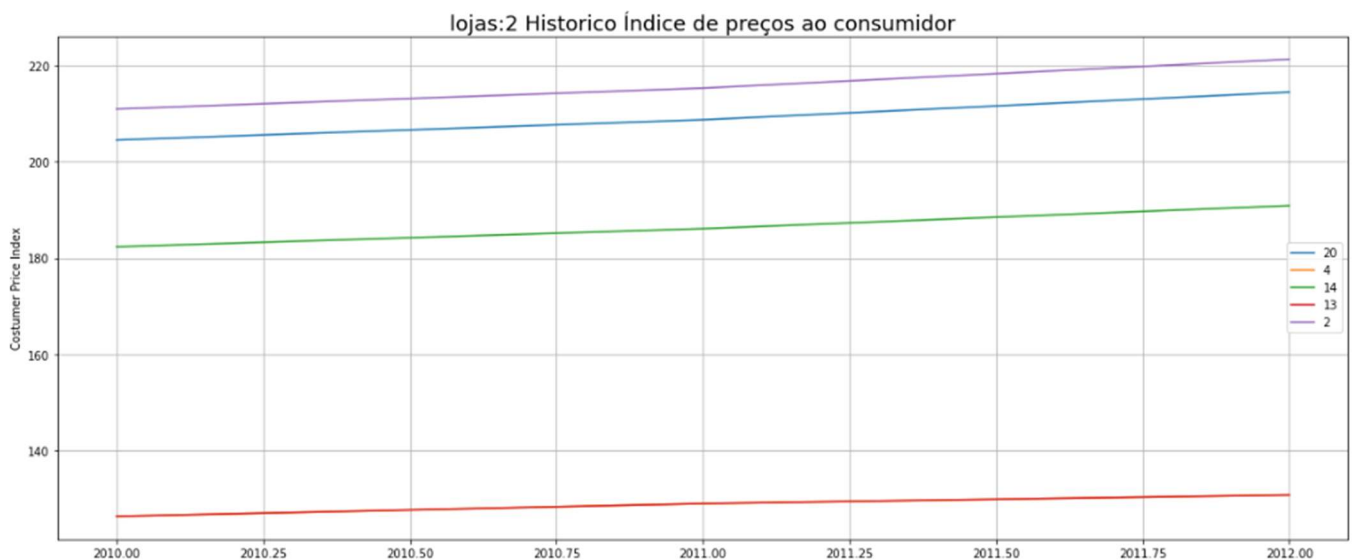
Sobre o Índice de preços ao Consumidor(*Constumer Price Index* - CPI):

```
[31] # Análise do histórico do índice de preços ao consumidor(CPI) (Lojas 20 , 4, 14, 13, 2 )

for x in sales_bigger:

    stores_bigger_sales = df.query('Store == @x')
    weekly_sales = stores_bigger_sales['CPI'].groupby(df['Year']).mean()
    weekly_sales.plot(figsize=(20,8));
    plt.title('lojas:%i Historico Índice de preços ao consumidor'%x, fontsize=18);
    plt.ylabel('Costumer Price Index', fontsize=10);
    plt.xlabel('')
    plt.grid(True)
    plt.legend(sales_bigger)

plt.show()
```



Quanto ao índice de preços ao consumidor, é possível verificar que as lojas 4 e 13 apresentam os menores valores, estando praticamente sobrepostas no gráfico.

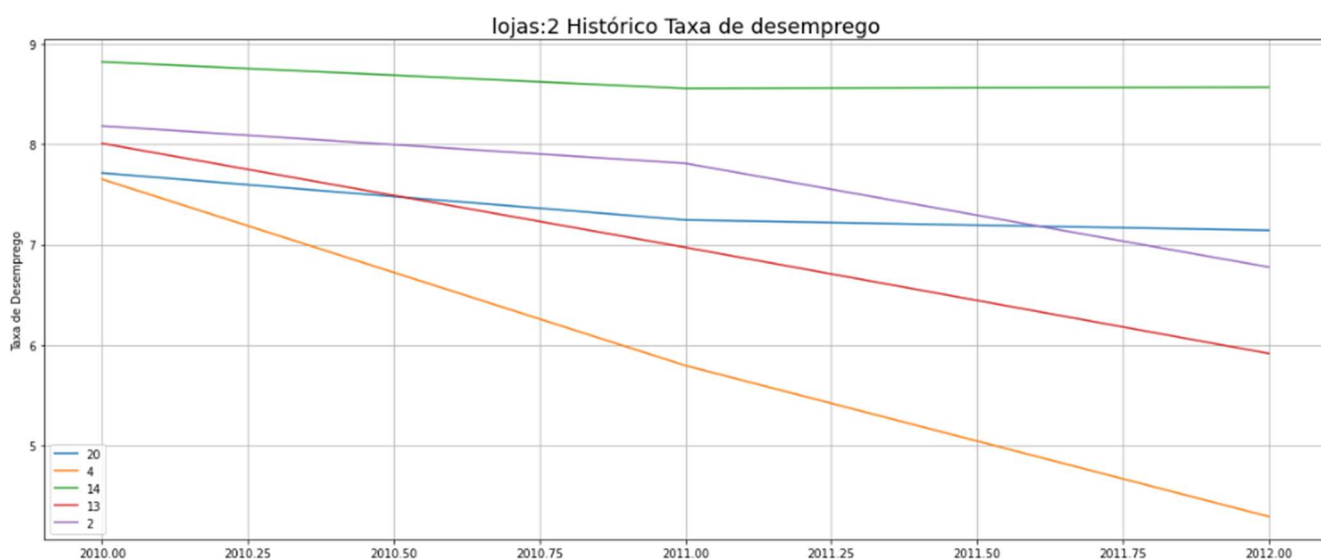
Por fim, quanto a taxa de desemprego:

```
[30] # Análise do histórico da Taxa de Desemprego (Lojas 20 , 4, 14, 13, 2 )

for x in sales_bigger:

    stores_bigger_sales = df.query('Store == @x')
    weekly_sales = stores_bigger_sales['Unemployment'].groupby(df['Year']).mean()
    weekly_sales.plot(figsize=(20,8));
    plt.title('lojas:%i Histórico Taxa de desemprego'%x, fontsize=18);
    plt.ylabel('Taxa de Desemprego', fontsize=10);
    plt.xlabel('')
    plt.grid(True)
    plt.legend(sales_bigger)

plt.show()
```



Analisando o histórico da Taxa de Desemprego entre as 5(cinco) lojas elencadas ao longo de todo o período, percebe-se que todas apresentaram queda na referida taxa, não obstante, a Loja 4 demonstra não só o menor valor, mas também a maior queda nesta taxa.

A diminuição de desempregados pode ser considerado fator concretamente relevante para a melhora na qualidade de vida e potencial financeiro da população, o que resulta consequentemente no aumento do poder de compra das pessoas naquela região.

6. Conclusão

Isto posto, diante dos fatores analisados, levando em consideração as métricas estabelecidas, a **Loja 4** e a **Loja 20** se apresentam entre os maiores faturamentos médios gerais. Entretanto é possível perceber, no período analisado, que a enquanto a **Loja 20** apresenta uma leve queda no faturamento médio, a **Loja 4** apresenta crescimento recorrente ao passar dos anos.

A Loja 13, ainda que tenha apresentado crescimento moderado no faturamento médio do período examinado, não supera os indicadores da Loja 4.

Insta frisar, que a **Loja 4** apresentou ainda uma vertiginosa e constante queda na taxa de desemprego, o que indica o aumento do poder de compra daquela região, além de um baixo índice de Preços ao Consumidor (CPI), comparada a outras Lojas no mesmo período.

A análise da variação do valor do combustível, assim como a variação de temperatura, não apresentou consistente importância em detrimento dos demais fatores verificados, não apresentando qualquer alteração considerável no faturamento e crescimento entre as lojas examinadas.

Desta forma, em face da profunda análise dos dados supra, indicamos a **LOJA 4** para investimento e expansão, por se apresentar com a maior probabilidade de lucratividade e retorno do referido investimento.

Destaque-se ainda, que de forma geral, as lojas analisadas apresentaram acentuado e recorrente aumento das vendas entre os meses de novembro e dezembro, reconhecidamente período dos feriados de *Thanksgiving* e *Christmas*, motivo pelo qual, indicamos que as obras de expansão sejam realizadas no lapso temporal anterior ou posterior ao referido período, com o fim de preservar o alto faturamento desta época do ano.

D4G11

[LINK COLAB:](#)

<https://colab.research.google.com/drive/1teOeMjwTGzMeFqfq2QkrkmYkCup9Dxaq?usp=sharing>